Resource Sharing Between Non-direct D2D Users and Cellular Users

Jiaqi Su[1], Wenbin Zhang[1,2](✉), Liyan Zhang[1], Jinwei Huang[1], and Shaochuan Wu[1]

[1] Harbin Institute of Technology, Harbin 150001, People's Republic of China
`zwbgxy1973@hit.edu.cn`
[2] Science and Technology on Communication Networks Laboratory, Shijiazhuang 050050, Hebei, People's Republic of China

Abstract. In recent years, mobile communication technology has developed rapidly and people have higher requirements for the communication quality. Addressing the problem of low communication quality in areas such as the edge of LTE-A cells, we introduce relay technology to expand the coverage area and improve the communication quality in hotspots. D2D technology can use the increasingly scarce spectrum resources efficiently and improve the utilization efficiency of communication system throughput and frequency band. Using layer 3 relay, we coexist with D2D direct users, D2D non-direct users, and cellular users to construct a communication network, and study the influence of different mode configurations on the throughput.

This paper considers the combination of relay technology and D2D technology in the LTE-A cell to study and analyze the communication performance of the system. Based on the LTE-A relay cellular network system architecture and link model and the communication mode of D2D technology, we combine the D2D undirected model with the LTE-A relay cellular network to establish a D2D system model, analyze the communication process, then construct a MINLP optimization problem with system throughput as the objective function and use the method of Lagrange dual to solve it. Through theoretical derivation and simulation analysis, the throughput of the LTE-A cell constructed by the combination of D2D non-direct mode and relay technology is obtained.

Keywords: LTE-Advanced cellular networks · Relay technology · D2D technology · Convex optimization

1 Introduction

1.1 LTE-A Relay Cellular Network

The introduction of relay technology in LTE-A allows relay nodes to expand the coverage area of the cell, and provide service signal coverage for areas with

---

Supported by National Nature Science Foundation of China (NSFC) under Grant 62071148.

Q. Guo et al. (Eds.): WiSATS 2021, LNICST 410, pp. 654–672, 2022.
https://doi.org/10.1007/978-3-030-93398-2_57

severe shadow fading in the cell, areas that cannot be covered by the signal due to the occlusion of buildings, and indoor and hotspot areas [1]. The specific implementation of the relay technology is to add some relay nodes or relay stations on the basis of the existing sites to make the distribution of the sites or antennas more dense. The user terminals use the intermediate access point, that is relay to access the network so as to obtain bandwidth services [2,3]. The use of relay nodes can reduce the loss of wireless links in free space, increase the signal-to-noise ratio, and achieve the purpose of increasing the information rate of users in edge areas. The newly added relay nodes and the existing parent base stations are connected wirelessly. When the downlink data is to be transmitted, the data to be transmitted firstly arrives at the parent base station, then it is transmitted from the parent base station to the relay node, and finally transmitted to the terminal, and vice versa is the uplink data transmission process. The data transmission method via the relay node is equivalent to narrowing the distance between the antenna and the terminal, which effectively improves the link quality in the terminal user information transmission process, and therefore can improve the system information rate and spectrum efficiency. In terms of network construction complexity and cost, the complexity of building a relay device is much lower than that of building a base station. The use of relay technology to achieve coverage in a cell can greatly reduce costs [4–7].

The LTE-A system was originally a traditional single-hop cellular network. After the introduction of relays, it becomes a relay cellular network, which is two-hop or multi-hop. This paper only studies the two-hop relay cellular network. Figure 1 shows the network architecture of the LTE-A relay cellular system under R10. Compared with the original LTE-A system architecture, it adds other interfaces such as RS. The user UE is connected to the RS through the interface Uu, and the RN is connected to the corresponding main base station DeNB through the interface Un. In the relay node protocol architecture, the DeNB is responsible for all interfaces X2 and interfaces S1 in the relay cellular network. After the introduction of relay nodes in the network, more communication links will inevitably be produced. According to the different service objects, the links in the cell where the relay exists are divided into three types, namely, the access link, the relay link, and the direct link [8–10]. In Fig. 2, there are a base station eNB, a relay node RN, and a user UE in the cell. The relay communicates with its covered users through the access link (L4), the relay communicates with the base station through the relay link (L3), and the base station communicates directly with the user through the direct link (L1, L2) [11–13].

1.2 D2D Communication Technology

How D2D Communication Technology Works. D2D communication refers to end-to-end direct communication technology under the control of cell base stations. D2D technology can use frequency band resources more efficiently, to a certain extent can alleviate the problem of shortage of communication spectrum resources, and can improve the throughput and spectrum utilization efficiency of the communication system. Figure 3 shows the D2D communication link model.

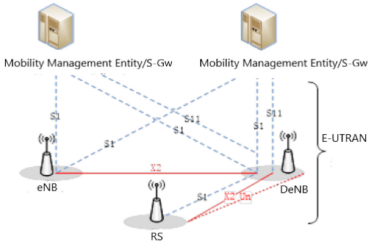

Fig. 1. Network architecture of relay cellular network

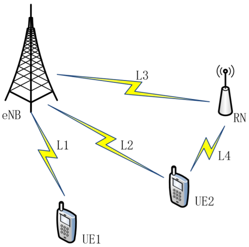

Fig. 2. Link model of relay cellular network

In the figure, the dotted line represents the control link, and the solid line represents the data link. The communication process is as follows: firstly, the user sends a request to establish D2D communication to the base station, and then the base station analyzes the request information to allocate a certain channel resource to the D2D user, and the user can use it directly to start the data communication process. Compared with all communication processes that have to be forwarded by the base station, the D2D communication method can effectively share the base station traffic and improve the system throughput under the condition of limited spectrum resources.

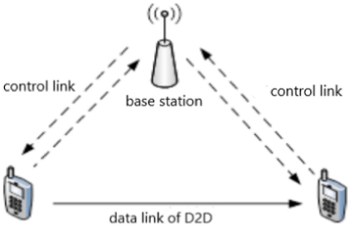

Fig. 3. D2D communication link model

D2D Communication User Working Mode. The introduction of D2D technology in a cell can increase the information rate at the edge of the cell or at the blind spot in areas that cannot be covered. D2D user clusters which contain several D2D users are mainly distributed in the edge area of a cell. Assuming that two users need to transmit information, then these two users form a D2D user pair. There are three ways for D2D user pairs to communicate: dedicated mode, multiplexing mode, and cellular mode.

In the dedicated mode, the system will allocate dedicated channel resources for D2D users to use, and the communication process of D2D users will not cause interference to the communication process of cellular users. Considering that the dedicated mode has relatively high requirements for frequency resources and will result in lower frequency utilization of the system, this mode is rarely used in actual situations.

D2D users in multiplexing mode need to share frequency (or time slot) resources with other cellular users in the cell. In this mode, there will be mutual interference between the communication process of D2D users and the communication process of cellular users, but the advantage is that it can improve spectrum utilization efficiency and increase system capacity [14,15]. Therefore, when a D2D user pair shares cellular user frequency resources, it is necessary to weigh the interference cost and communication performance, and to ensure both the communication quality of the cellular users and the D2D users, and can't save channel resources or increase the utilization rate at the expense of affecting communication performance. Subsequent chapters will analyze the interference situation in multiplexing mode in detail.

The dedicated mode and the multiplexing mode can be interchanged in some cases: when the cell has high requirements for communication reliability and the traffic volume is small, D2D users will be allowed to use the dedicated mode to communicate [16,17]. When the cell traffic is large, the system will allow D2D users to preferentially use the multiplexing mode for communication in order to save spectrum resources.

Cellular mode means that D2D users need to be relayed through the base station based on the multiplexing mode. At this time, D2D communication requires both uplink and downlink communication link resources.

Interference Analysis of D2D Communication in Multiplexing Mode. The following analyzes the interference situation when D2D users share frequency resources with other cellular users in the cell, which means the D2D users are in the multiplexing mode. Cellular user communication is divided into two communication links, namely Uplink (UL) and Downlink (DL). Uplink (UL) refers to the communication from the user terminal to the base station, while downlink (DL) refers to the communication from the base station to the user terminal. Therefore, the ways in which D2D users multiplex cellular user frequency resources can be divided into two types: multiplexing cellular user uplink and multiplexing cellular user downlink. The following will be detailed analysis of these two multiplexing situations.

Figure 4 shows the link diagram when multiplexing uplink resources, where UE1 and UE2 are a D2D user pair. UE1 is the sender, UE2 is the receiver, and UE3 is a cellular user. The interference situation is that D2D users will cause interference to the base station when communicating from the sender to the receiving end, and communications from cellular users to the base station will also cause interference to the receiving end of the D2D user. Figure 5 is the link diagram when multiplexing downlink resources. The user distribution is the same as Fig. 4. The interference situation is that D2D users cause interference to cellular users when communicating from the sender to the receiver. Also, communication from base station to cellular user will cause interference to the receiving end of D2D users. Comparing the two cases, the interference caused to the receiving end of D2D users both exists. In addition, the multiplexing of uplink frequency resources causes interference to the base station, and the multiplexing of downlink frequency resources causes interference to many cellular users. But it is more difficult to control the interference items when multiplexing the downlink. Moreover, it is mentioned in the literature that the resource utilization efficiency of the uplink communication link will be lower than the resource utilization efficiency of the downlink communication link. Therefore, through the above analysis, this paper will consider the case of D2D users multiplexing the uplink communication links of cellular users.

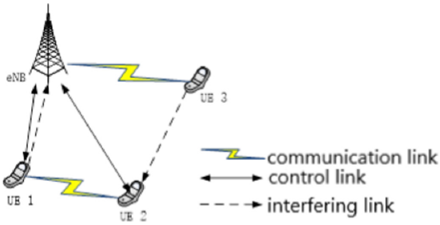

Fig. 4. D2D multiplexing cellular uplink

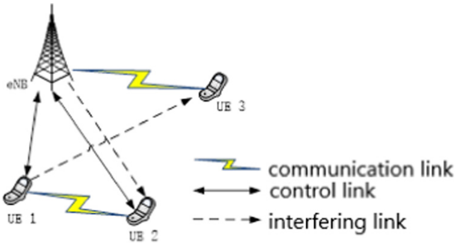

Fig. 5. D2D multiplexing cellular downlink

The rest of the paper is organized as follows. Section 2 briefly introduces the system model of D2D non-directed system. Section 3 analyses the formulation of the problem. The simulation results is shown in Sect. 4. Finally, conclusions are drawn in Sect. 5.

Notation: boldface uppercase, boldface lowercase, and lowercase letter $\mathbf{A}$, $\mathbf{a}$, a denote a matrix, vector, and scalar variable, respectively. $\mathbf{A}^T$, $\mathbf{A}^*$and $\mathbf{A}^{-1}$ represent the transpose, the conjugate transpose and the inverse of a matrix, respectively. $\|$ denotes absolute value, *tr()* denotes trace of a matrix, *diag* denotes diagnoal.

2 System Model

2.1 D2D Non-directed System Model

As shown in Fig. 6, the system model diagram contains both D2D non-direct users and cellular users. The same as in Sect. 1.2, the communication process is divided into two time slots, and all relays communicate synchronously in these two time slots.

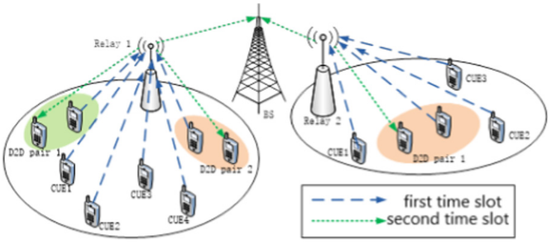

Fig. 6. A system model in which D2D non-direct users and cellular users coexist

The following analyzes the communication process of each user in turn. There are two communication modes for users under the coverage of each relay, namely:

1. The transmitter of the D2D non-direct user pair communicates with the receiver through relay and forwarding, including the communication between the transmitter of the D2D non-direct user and the relay, and the communication between the relay and the receiver of the D2D non-direct user.
2. Cellular users communication, including the communication between cellular users and relays, and the communication between relays and base stations.

It is assumed that the communication between the relays and base stations uses orthogonal channels, so there will be no interference between them.

The communication process in time slot one is from terminals to relays, and the communication process in time slot two includes from relays to terminals or relays to base stations. For cellular users, only the uplink is analyzed which includes two time slots. The first time slot is from the terminal to the relay, and the second time slot is from the relay to the base station. The communication process of each time slot will only be interfered by other communication processes in the own time slot.

Assume that set M represents all cellular users, and set D^c represents all non-direct D2D user pairs. The system bandwidth is divided into N resource blocks (RB), whose set is $N = \{1, 2, \cdots, |N|\}$. The usable resource block set in each relay is N, and the bandwidth of each resource block is represented by B_{RB}. The relay set is represented by $L = \{1, 2, \cdots |L|\}$, and $U_l, \forall l \in L$ is the set of all users who carry out the communication process assisted by relay l. The set of cellular users under the coverage of relay l is $M_l = M \cap U_l$, and the set of D2D non-direct users under the coverage of relay l is $D_l^c = D^c \cap U_l$. Therefore, the following formulas are established: $U_l = D_l^c \cup M_l, \forall l \in L \cup_l U_l = \{D^c \cup M\}$ $\cap_l U_l = \varphi, \forall l \in L$.

According to the Shannon channel capacity formula: $R = B\log_2(1 + S/N)$. The information rate expression of each link in the communication process of the system model can be listed separately.

(1) The communication between the transmitter of D2D non-direct users or cellular users and relay: the communication process from the cellular users under the coverage of the relay and the D2D non-direct users to the relay is completed in time slot 1, and both communication processes will be affected by interference from D2D non-direct users and cellular users under other relays. According to the characteristics that similar communication process has similar interference received, the unit power SINR of the communication process is expressed as follows, that is, for the transmitter $u_l \in D_l^c$ of the D2D non-direct user, the cellular user $u_l \in M_l$, there is:

$$\gamma_{u_l,l,1}^{(n)} = \frac{h_{u_l,l}^{(n)}}{\sum\limits_{\substack{u_j \in D_j^c \cup M_j \\ j \neq l, j \in L}} Q_{u_j,j}^{(n)} \cdot g_{u_j,l}^{(n)} + \sigma^2} \tag{1}$$

s.t.

$$\sum_{u_l \in U_l} y_{u_l}^{(n)} \leq 1, \quad \forall n \in N \tag{2}$$

Where $Q_{a,b}^{(n)}$, $h_{a,b}^{(n)}$ and $g_{a,b}^{(n)}$ are the transmission power, the channel coefficient of the communication link, and the interference link gain in the communication process from the transmitting end a to the receiving end b on the resource block n respectively. $\sigma^2 = N_0 B_{RB}$, $B_{RB}$is the resource block bandwidth, and N_0 is the power spectral density of the thermal noise.

(2) The communication between the receiver of D2D non-direct users pair and relay l: the communication between the receiver of D2D non-direct users pair and relay l is carried out in time slot 2. Consider that all communication processes in the second time slot will be affected by interference in the communication process between other relay j and the receiving end of

the D2D non-direct user in its own coverage area and interference in the communication process between other relay j and the base station. Assuming that within a certain time interval, only one relay-base station link is in the working state in the second time slot for transmitting cellular user data messages to the base station, while other relay-base station links are in the dormant state. Therefore, the unit power SINR of the communication between the receiver of D2D non-direct users pair and relay l on the resource block n is expressed as the following formula, that is, for $u_l \in D_l^c$, there is:

$$\gamma_{l,u_l,2}^{(n)} = \frac{h_{l,u_l}^{(n)}}{\sum\limits_{\substack{\forall u_j \in D_j^c \cup M_j \\ j \neq l, j \in L}} Q_{j,u_j}^{(n)} \cdot g_{j,u_l}^{(n)} + \sigma^2} \tag{3}$$

(3) The communication between relay l and base station: the communication between relay l and base station means that relay l sends information from cellular users in its coverage area to the base station through resource block n and the base station forwards it in the second time slot, so as to achieve communication with users in other cells. Assuming that within a certain time interval, only one relay-base station link is in the working state in the second time slot for transmitting cellular user data messages to the base station, and other relay-base station links are in the dormant state. Therefore, in this process, the base station will also be interfered by the communication process between other relay j and the D2D non-direct user receiving end under its own coverage on resource block n. Therefore, for $u_l \in M_l$, the SINR per unit power is:

$$\gamma_{l,u_l,2}^{(n)} = \frac{h_{l,eNB}^{(n)}}{\sum\limits_{\substack{\forall u_j \in D_j^c \\ j \neq l, j \in L}} Q_{j,u_j}^{(n)} \cdot g_{j,eNB}^{(n)} + \sigma^2} \tag{4}$$

In summary, for the user u_l under the coverage of relay l, the total information rate is composed of two parts: the information rate in the first time slot is $R_{u_l,l}^{(n)} = B_{RB}\log_2\left(1 + Q_{u_l,l}^{(n)}\gamma_{u_l,l,1}^{(n)}\right), u_l \in D_l^c \cap M_l$, and the information rate in the second time slot is $R_{l,u_l}^{(n)} = B_{RB}\log_2\left(1 + Q_{l,u_l}^{(n)}\gamma_{l,u_l,2}^{(n)}\right), u_l \in D_l^c \cap M_l$. Therefore, for user u_l under the coverage of relay l, the total information rate on resource block n is denoted as $R_{u_l}^{(n)} = \frac{1}{2}\min\left\{R_{u_l,l}^{(n)}, R_{l,u_l}^{(n)}\right\}$.

3 Problem Formulation

3.1 Problem Modeling

This section discusses that under the premise that each transmitting node of the D2D link satisfies the transmission power constraint and the interference to the cellular network is less than the interference threshold, and combined with the relay and forwarding technology, how to allocate resources reasonably for cellular users and D2D non-direct users to maximize the throughput of the system.

The system capacity is maximized through resource block allocation and power allocation. For a communication process completed through two time slots, the total communication rate is determined by the smaller of the two time slot communication rates. Set the maximum value of user transmission power to $Q_{u_l}^{\max}$, and the maximum value of relay transmission power to $Q_l^{\max}$. In order to express that each resource block RB can only be used by one user under the relay, a resource block allocation factor of $y_{u_l}^{(n)}$ is introduced, and $y_{u_l}^{(n)} \in \{0,1\}$ is a binary integer variable, which means when $y_{u_l}^{(n)} = 1$, it means that resource block n is allocated to user u_l for use; otherwise, $y_{u_l}^{(n)} = 0$. For all users u_l under relay l, the total information rate is $R_{u_l} = \sum_{u_l \in U_l} \sum_{n=1}^{N} y_{u_l}^{(n)} R_{u_l}^{(n)}$. The user's QoS requirement is represented by R_{QoS}. Considering that the same resource block will be occupied by the relay in two time slots, this optimization problem can be described as:

$$\max_{y_{u_l}^{(n)}, Q_{u_l,l}^{(n)}, Q_{l,u_l}^{(n)}} \sum_{l \in L} \sum_{u_l \in U_l} \sum_{n=1}^{N} y_{u_l}^{(n)} R_{u_l}^{(n)} \tag{5}$$

s.t.

$$\sum_{u_l \in U_l} y_{u_l}^{(n)} \leq 1, \quad \forall n \in N \tag{6}$$

$$\sum_{n=1}^{N} y_{u_l}^{(n)} Q_{u_l,l}^{(n)} \leq Q_{u_l}^{\max}, \quad \forall u_l \in U_l \tag{7}$$

$$\sum_{u_l \in U_l} \sum_{n=1}^{N} y_{u_l}^{(n)} Q_{l,u_l}^{(n)} \leq Q_l^{\max}, \ \forall l \in L \tag{8}$$

$$\sum_{u_l \in U_l} y_{u_l}^{(n)} Q_{u_l,l}^{(n)} g_{u_l,l^*}^{(n)} \leq I_{th,1}^{(n)}, \ \forall n \in N \tag{9}$$

$$\sum_{u_l \in D_l^c} y_{u_l}^{(n)} Q_{l,u_l}^{(n)} g_{l,u_l^*}^{(n)} \leq I_{th,2}^{(n)}, \ \forall n \in N \tag{10}$$

$$R_{u_l} \geq R_{QoS}, \quad \forall u_l \in U_l \tag{11}$$

$$Q_{u_l,l}^{(n)} \geq 0, \quad Q_{l,u_l}^{(n)} \geq 0, \quad \forall n \in N, \ u_l \in U_l \tag{12}$$

Where constraint (6) is the condition that each indicator coefficient needs to meet, and each resource block RB can only be allocated to one user under each relay. Constraint (7) is that the transmission power of each user cannot exceed the maximum power limit. Constraint (8) means that the transmit power of each relay station cannot exceed the maximum power limit. Constraint (9) means that the interference from users under other relays needs to meet the interference threshold. Constraint (10) means that the interference from other relay stations need to meet the threshold. Constraint (11) ensures that the system meets the minimum QoS requirements. Constraint (12) means that each transmission power is non-negative.

For all users u_l on resource block n, the information rate is:

$$\begin{aligned} y_{u_l}^{(n)} R_{u_l}^{(n)} &= \tfrac{1}{2} y_{u_l}^{(n)} \min \left\{ R_{u_l,l}^{(n)}, R_{l,u_l}^{(n)} \right\} \\ &= \tfrac{1}{2} y_{u_l}^{(n)} \min \left\{ B_{RB} \log_2 \left(1 + Q_{u_l,l}^{(n)} \gamma_{u_l,l,1}^{(n)}\right), B_{RB} \log_2 \left(1 + Q_{l,u_l}^{(n)} \gamma_{l,u_l,2}^{(n)}\right) \right\} \end{aligned} \tag{13}$$

For D2D non-direct users or cellular users, the unit power SINR in the communication process in time slot 1 is:

$$\begin{gathered} \gamma_{u_l,l,1}^{(n)} = \frac{h_{u_l,l}^{(n)}}{I_{u_l,l,1}^{(n)} + \sigma^2} \\ I_{u_l,l,1}^{(n)} = \sum_{\substack{u_j \in D_j^c \cup M_j \\ j \neq l, j \in L}} y_{u_j}^{(n)} Q_{u_j,j}^{(n)} \cdot g_{u_j,l}^{(n)} \end{gathered} \tag{14}$$

Where $I_{u_l,l,1}^{(n)}$ is the interference item suffered by user u_l on resource block n in the first time slot. For D2D non-direct users and cellular users, the unit power SINR in the communication process in time slot 2 is:

$$\gamma_{l,u_l,2}^{(n)} = \begin{cases} \frac{h_{l,u_l}^{(n)}}{I_{l,u_l,2}^{(n)} + \sigma^2}, & u_l \in D_l^c \\ \frac{h_{l,eNB}^{(n)}}{I_{l,u_l,2}^{(n)} + \sigma^2}, & u_l \in M_l \end{cases} \tag{15}$$

Where $I_{l,u_l,2}^{(n)}$ is the interference item suffered by user u_l on resource block n in the second time slot.

$$I_{l,u_l,2}^{(n)} = \begin{cases} \sum\limits_{\substack{\forall u_j \in D_j^c \cup M_j \\ j \neq l, j \in L}} y_{u_j}^{(n)} Q_{j,u_j}^{(n)} \cdot g_{j,u_l}^{(n)}, & u_l \in D_l^c \\ \sum\limits_{\substack{\forall u_j \in D_j^c \\ j \neq l, j \in L}} y_{u_j}^{(n)} Q_{j,u_j}^{(n)} \cdot g_{j,eNB}^{(n)}, & u_l \in M_l \end{cases} \tag{16}$$

When $Q_{u_l,l}^{(n)} \gamma_{u_l,l,1}^{(n)} = Q_{l,u_l}^{(n)} \gamma_{l,u_l,2}^{(n)}$ is established, the information rate $R_{u_l}^{(n)}$ of all users u_l communicating on resource block n can reach the maximum value. At this time, $Q_{l,u_l}^{(n)}$ in the second time slot can be represented by the power in the first time slot, that is, $Q_{l,u_l}^{(n)} = \frac{\gamma_{u_l,l,1}^{(n)}}{\gamma_{l,u_l,2}^{(n)}} Q_{u_l,l}^{(n)}$. Therefore, the information rate $R_{u_l}^{(n)}$ of user u_l on resource block n can be rewritten as:

$$R_{u_l}^{(n)} = \frac{1}{2} B_{RB} \log_2 \left(1 + Q_{u_l,l}^{(n)} \gamma_{u_l,l,1}^{(n)}\right) \tag{17}$$

The optimization problem (5) contains both continuous variables and binary integer variables, and the objective function is also nonlinear, which is a MINLP problem. In order to simplify the problem, first we need to relax the resource block allocation factor $y_{u_l}^{(n)}$ to a continuous variable, that is, $y_{u_l}^{(n)} \in (0,1]$. $y_{u_l}^{(n)}$ represents the proportion of time that is allocated to the user u_l by the resource block n, and the restriction condition $\sum\limits_{u_l \in U_l} y_{u_l}^{(n)} \le 1$, $\forall n \in N$ mentioned above is still met. We also introduce power allocation variable $T_{u_l,l}^{(n)} = y_{u_l}^{(n)} Q_{u_l,l}^{(n)}$ to represent the actual transmit power of user u_l when communicating on resource block n, then the optimization problem after relaxation and adjustment can be expressed as:

$$\max_{y_{u_l}^{(n)}, T_{u_l,l}^{(n)}, \mu_{u_l}^{(n)}} \sum_{l \in L} \sum_{u_l \in U_l} \sum_{n=1}^{N} \frac{1}{2} y_{u_l}^{(n)} B_{RB} \log_2 \left(1 + \frac{T_{u_l,l}^{(n)} h_{u_l,l}^{(n)}}{y_{u_l}^{(n)} \mu_{u_l}^{(n)}} \right) \tag{18}$$

s.t.

$$\sum_{u_l \in U_l} y_{u_l}^{(n)} \le 1, \quad \forall n \in N \tag{19}$$

$$\sum_{n=1}^{N} T_{u_l,l}^{(n)} \le Q_{u_l}^{\max}, \quad \forall u_l \in U_l \tag{20}$$

$$\sum_{u_l \in U_l} \sum_{n=1}^{N} \frac{\gamma_{u_l,l,1}^{(n)}}{\gamma_{l,u_l,2}^{(n)}} T_{u_l,l}^{(n)} \le Q_l^{\max} \tag{21}$$

$$\sum_{u_l \in U_l} T_{u_l,l}^{(n)} g_{u_l,l^*}^{(n)} \le I_{th,1}^{(n)}, \quad \forall n \in N \tag{22}$$

$$\sum_{u_l \in D_l^c} \frac{\gamma_{u_l,l,1}^{(n)}}{\gamma_{l,u_l,2}^{(n)}} T_{u_l,l}^{(n)} g_{l,u_l^*}^{(n)} \le I_{th,2}^{(n)}\ , \ \forall n \in N \tag{23}$$

$$\sum_{n=1}^{N} \frac{1}{2} y_{u_l}^{(n)} B_{RB} \log_2 \left(1 + \frac{T_{u_l,l}^{(n)} h_{u_l,l}^{(n)}}{y_{u_l}^{(n)} \mu_{u_l}^{(n)}} \right) \ge R_{QoS}\, , \ \forall u_l \in U_l \tag{24}$$

$$T_{u_l,l}^{(n)} \ge 0, \quad \forall n \in N, \ u_l \in U_l \tag{25}$$

$$I_{u_l,l}^{(n)} + \sigma^2 \le \mu_{u_l}^{(n)}, \ \forall n \in N, \ u_l \in U_l \tag{26}$$

Where $\mu_{u_l}^{(n)}$ is an auxiliary variable, and there is $I_{u_l,l}^{(n)} = \max \left\{ I_{u_l,l,1}^{(n)}, I_{l,u_l,2}^{(n)} \right\}$. As the number of resource blocks increases, when the number is extremely large, the dual spacing of the optimization problem that satisfies the $y_{u_l}^{(n)}$ time allocation condition can be ignored. The optimization problem studied in this paper satisfies the time allocation condition, so the relaxed optimization problem has an asymptotic optimal solution. It can be seen from the above formula that the constraint (24) is convex, and other constraints are linear. If the objective function is a concave function, then the optimization problem (18) is a convex problem and there is an optimal solution. The following first proves that the objective optimization function is a concave function.

Define function $\Re\left(T_{u_l,l}^{(n)}\right) = -y_{u_l}^{(n)} B_{RB} \log_2\left(1+\frac{T_{u_l,l}^{(n)} h_{u_l,l}^{(n)}}{y_{u_l}^{(n)} \mu_{u_l}^{(n)}}\right)$, and find the Hessian matrix H of power distribution variable $T_{u_l,l}^{(n)}$ for the function is:

$$H = \left| \frac{\partial^2 \Re\left(T_{u_l,l}^{(n)}\right)}{\partial {T_{u_l,l}^{(n)}}^2} \right| = \left| \frac{y_{u_l}^{(n)} B_{RB} {h_{u_l,l}^{(n)}}^2}{\ln 2\left(y_{u_l}^{(n)} \mu_{u_l}^{(n)} + T_{u_l,l}^{(n)} h_{u_l,l}^{(n)}\right)^2} \right| \tag{27}$$

H is a first-order matrix, so the eigenvalue is $\tilde{\lambda} = \frac{y_{u_l}^{(n)} B_{RB} {h_{u_l,u_l}^{(n)}}^2}{\ln 2\left(y_{u_l}^{(n)} \mu_{u_l}^{(n)} + T_{u_l,l}^{(n)} h_{u_l,l}^{(n)}\right)^2}$. Because of $\tilde{\lambda} > 0$, $\Re\left(T_{u_l,l}^{(n)}\right)$ is convex, and the objective function (18) is concave. Therefore, the optimization problem is a convex problem, and there is an optimal solution. The KKT condition in the convex optimization theory can be used to solve this problem.

3.2 Power Distribution Method Based on Lagrange Multiplier Method

In Sect. 3.1, it has been proved that the optimization problem (18) is a convex problem. The Lagrange multiplier method is used to solve the problem, and the Lagrange multiplier of the constraints (19) to (26) of formula (18) are $\delta_n, \xi_{u_l}, \upsilon_l, \psi_n, \varepsilon_n, \lambda_{u_l}, \rho_{u_l}^{(n)}$ respectively, then the Lagrange function is:

$$\begin{aligned}
L = &-\sum_{l\in L}\sum_{u_l\in U_l}\sum_{n=1}^{N} \tfrac{1}{2} y_{u_l}^{(n)} B_{RB} \log_2\left(1+\tfrac{T_{u_l,l}^{(n)} h_{u_l,l}^{(n)}}{y_{u_l}^{(n)} \mu_{u_l}^{(n)}}\right) \\
&+ \sum_{n=1}^{N} \delta_n \left(\sum_{u_l\in U_l} y_{u_l}^{(n)} - 1\right) + \sum_{u_l\in U_l} \xi_{u_l} \left(\sum_{n=1}^{N} T_{u_l,l}^{(n)} - Q_{u_l}^{\max}\right) \\
&+ \upsilon_l \left(\sum_{u_l\in U_l}\sum_{n=1}^{N} \tfrac{\gamma_{u_l,l,1}^{(n)}}{\gamma_{l,u_l,2}^{(n)}} T_{u_l,l}^{(n)} - Q_l^{\max}\right) + \sum_{n=1}^{N} \psi_n \left(\sum_{u_l\in U_l} T_{u_l,l}^{(n)} g_{u_l,l^*}^{(n)} - I_{th,1}^{(n)}\right) \\
&+ \sum_{n=1}^{N} \varepsilon_n \left(\sum_{u_l\in D_l^c} \tfrac{\gamma_{u_l,l,1}^{(n)}}{\gamma_{l,u_l,2}^{(n)}} T_{u_l,l}^{(n)} g_{l,u_l^*}^{(n)} - I_{th,2}^{(n)}\right) + \sum_{u_l\in U_l}\sum_{n=1}^{N} \rho_{u_l}^{(n)} \left(I_{u_l,l}^{(n)} + \sigma^2 - \mu_{u_l}^{(n)}\right) \\
&+ \sum_{u_l\in U_l} \lambda_{u_l} \left(R_{QoS} - \sum_{n=1}^{N} \tfrac{1}{2} y_{u_l}^{(n)} B_{RB} \log_2\left(1+\tfrac{T_{u_l,l}^{(n)} h_{u_l,l}^{(n)}}{y_{u_l}^{(n)} \mu_{u_l}^{(n)}}\right)\right)
\end{aligned} \tag{28}$$

First, the derivative of the power distribution variable $T_{u_l,l}^{(n)}$ is obtained, and the following formula is obtained:

$$\begin{aligned}
\frac{\partial L}{\partial T_{u_l,l}^{(n)}} = &-\tfrac{1}{2\ln 2} y_{u_l}^{(n)} B_{RB} \frac{1}{1+\frac{T_{u_l,l}^{(n)} h_{u_l,l}^{(n)}}{y_{u_l}^{(n)} \mu_{u_l}^{(n)}}} \frac{h_{u_l,l}^{(n)}}{y_{u_l}^{(n)} \mu_{u_l}^{(n)}} + \xi_{u_l} + \upsilon_l \frac{\gamma_{u_l,l,1}^{(n)}}{\gamma_{l,u_l,2}^{(n)}} + \varepsilon_n \frac{\gamma_{u_l,l,1}^{(n)}}{\gamma_{l,u_l,2}^{(n)}} g_{l,u_l^*}^{(n)} \\
&+ \psi_n g_{u_l,l^*}^{(n)} - \tfrac{1}{2\ln 2} \lambda_{u_l} y_{u_l}^{(n)} B_{RB} \frac{1}{1+\frac{T_{u_l,l}^{(n)} h_{u_l,l}^{(n)}}{y_{u_l}^{(n)} \mu_{u_l}^{(n)}}} \frac{h_{u_l,l}^{(n)}}{y_{u_l}^{(n)} \mu_{u_l}^{(n)}}
\end{aligned} \tag{29}$$

According to the KKT condition in the convex optimization theory, let $\frac{\partial L}{\partial T_{u_l,l}^{(n)}} = 0$, wee can get the optimal solution of the transmission power $Q_{u_l,l}^{(n)}$ from user u_l to relay l as follows:

$$\begin{gathered} Q_{u_l,l}^{(n)^*} = \frac{T_{u_l,l}^{(n)^*}}{y_{u_l}^{(n)^*}} = \left[\Delta_{u_l,l}^{(n)} - \frac{\mu_{u_l}^{(n)}}{h_{u_l,l}^{(n)}}\right]^* \\ \Delta_{u_l,l}^{(n)} = \frac{(\lambda_{u_l}+1)\,B_{RB}}{2\ln 2\left(\xi_{u_l} + \upsilon_l \frac{\gamma_{u_l,l,1}^{(n)}}{\gamma_{l,u_l,2}^{(n)}} + \psi_n g_{u_l,l^*}^{(n)} + \varepsilon_n \frac{\gamma_{u_l,l,1}^{(n)}}{\gamma_{l,u_l,2}^{(n)}} g_{l,u_l^*}^{(n)}\right)} \end{gathered} \tag{30}$$

Where $[\xi]^+ = \max\{\xi, 0\}$. Derivation of the resource block allocation factor $y_{u_l}^{(n)}$ after relaxation, we get:

$$\begin{aligned} \frac{\partial L}{\partial y_{u_l}^{(n)}} = &-\tfrac{1}{2}B_{RB}\log_2\left(1 + \frac{T_{u_l,l}^{(n)}h_{u_l,l}^{(n)}}{y_{u_l}^{(n)}\mu_{u_l}^{(n)}}\right) - \tfrac{1}{2}\lambda_{u_l}B_{RB}\log_2\left(1 + \frac{T_{u_l,l}^{(n)}h_{u_l,l}^{(n)}}{y_{u_l}^{(n)}\mu_{u_l}^{(n)}}\right) + \delta_n \\ &- \tfrac{1}{2\ln 2}y_{u_l}^{(n)}B_{RB}\frac{1}{1+\frac{T_{u_l,l}^{(n)}h_{u_l,l}^{(n)}}{y_{u_l}^{(n)}\mu_{u_l}^{(n)}}}\frac{T_{u_l,l}^{(n)}h_{u_l,l}^{(n)}}{\mu_{u_l}^{(n)}}\left(-\frac{1}{y_{u_l}^{(n)2}}\right) \\ &- \tfrac{1}{2\ln 2}y_{u_l}^{(n)}\lambda_{u_l}B_{RB}\frac{1}{1+\frac{T_{u_l,l}^{(n)}h_{u_l,l}^{(n)}}{y_{u_l}^{(n)}\mu_{u_l}^{(n)}}}\frac{T_{u_l,l}^{(n)}h_{u_l,l}^{(n)}}{\mu_{u_l}^{(n)}}\left(-\frac{1}{y_{u_l}^{(n)2}}\right) \end{aligned} \tag{31}$$

According to the KKT condition in the convex optimization theory, let $\frac{\partial L}{\partial y_{u_l}^{(n)}} = 0$, we get:

$$\delta_n = \frac{1}{2}(1+\lambda_{u_l})\,B_{RB}\left(\log_2\left(1 + \frac{T_{u_l,l}^{(n)}h_{u_l,l}^{(n)}}{y_{u_l}^{(n)}\mu_{u_l}^{(n)}}\right) - \theta_{u_l,l}^{(n)}\right) \tag{32}$$

Where $\theta_{u_l,l}^{(n)} = \frac{T_{u_l,l}^{(n)}h_{u_l,l}^{(n)}}{\ln 2\left(y_{u_l}^{(n)}\mu_{u_l}^{(n)} + T_{u_l,l}^{(n)}h_{u_l,l}^{(n)}\right)}$. The actual value of δ_n is not calculated by formula (32), but updated by sub-gradient iteration method. In order to obtain the integer value of resource block allocation factor $y_{u_l}^{(n)}$, the threshold variable $\chi_{u_l}^{(n)} = \frac{1}{2}(1+\lambda_{u_l})\,B_{RB}\left(\log_2\left(1 + \frac{T_{u_l,l}^{(n)}h_{u_l,l}^{(n)}}{y_{u_l}^{(n)}\mu_{u_l}^{(n)}}\right) - \theta_{u_l,l}^{(n)}\right)$ needs to be set to obtain the discriminant of resource block allocation factor $y_{u_l}^{(n)}$ is as follows:

$$y_{u_l}^{(n)^*} = \begin{cases} 1, & \delta_n \le \chi_{u_l}^{(n)} \\ 0, & \delta_n > \chi_{u_l}^{(n)} \end{cases} \tag{33}$$

After obtaining the optimal solution of transmission power $Q_{u_l,l}^{(n)^*}$ and resource block allocation factor $y_{u_l}^{(n)^*}$ for communication from user u_l to relay l , the sub-gradient iteration method is used to update each Lagrange multiplier, and the Lagrange multiplier of the (t + 1)-th iteration is updated according to the following formulas, and the variable $\varLambda_{\alpha}^{(t)}$ is the step length in the first iteration, where $\varLambda_{\alpha}^{(t)} = a/\sqrt{t}$, and a is constant.

$$\delta_n\left(t+1\right) = \left[\delta_n\left(t\right) + \varLambda_{\delta_n}^{(t)}\left(\sum_{u_l \in U_l} y_{u_l}^{(n)} - 1\right)\right]^{+} \tag{34}$$

$$\xi_{u_l}\left(t+1\right) = \left[\xi_{u_l}\left(t\right) + \varLambda_{\xi_{u_l}}^{(t)}\left(\sum_{n=1}^{N} T_{u_l,l}^{(n)} - Q_{u_l}^{\max}\right)\right]^{+} \tag{35}$$

$$\upsilon_l\left(t+1\right) = \left[\upsilon_l\left(t\right) + \varLambda_{\upsilon_l}^{(t)}\left(\sum_{u_l \in U_l}\sum_{n=1}^{N} \frac{\gamma_{u_l,l,1}^{(n)}}{\gamma_{l,u_l,2}^{(n)}} T_{u_l,l}^{(n)} - Q_l^{\max}\right)\right]^{+} \tag{36}$$

$$\psi_n\left(t+1\right) = \left[\psi_n\left(t\right) + \varLambda_{\psi_n}^{(t)}\left(\sum_{u_l \in U_l} T_{u_l,l}^{(n)} g_{u_l,l^*}^{(n)} - I_{th,1}^{(n)}\right)\right]^{+} \tag{37}$$

$$\varepsilon_n\left(t+1\right) = \left[\varepsilon_n\left(t\right) + \varLambda_{\varepsilon_n}^{(t)}\left(\sum_{u_l \in D_l^c} \frac{\gamma_{u_l,l,1}^{(n)}}{\gamma_{l,u_l,2}^{(n)}} T_{u_l,l}^{(n)} g_{l,u_l^*}^{(n)} - I_{th,2}^{(n)}\right)\right]^{+} \tag{38}$$

$$\lambda_{u_l}\left(t+1\right) = \left[\lambda_{u_l}\left(t\right) + \varLambda_{\lambda_{u_l}}^{(t)}\left(R_{QoS} - \sum_{n=1}^{N} \frac{1}{2} y_{u_l}^{(n)} B_{RB} \log_2\left(1 + \frac{T_{u_l,l}^{(n)} h_{u_l,l}^{(n)}}{y_{u_l}^{(n)} \mu_{u_l}^{(n)}}\right)\right)\right]^{+} \tag{39}$$

$$\rho_{u_l}^{(n)}\left(t+1\right) = \left[\rho_{u_l}^{(n)}\left(t\right) + \varLambda_{\rho_{u_l}^{(n)}}^{(t)}\left(I_{u_l,l}^{(n)} + \sigma^2 - \mu_{u_l}^{(n)}\right)\right]^{+} \tag{40}$$

4 Simulation Results and Performance Analysis

Since each channel coefficient and interference link channel coefficient in the simulation contain random variables, for each resource block, the information rate distributed on it also has a certain randomness. In order to reduce the final randomness caused by randomness, the method of calculating the average value by multiple simulations is used to eliminate the influence of the randomness of the channel coefficient. After this process, we can obtain the fairness index curve of the information rate distribution on each resource block. This paper uses the RajJain fairness index to determine the fairness of the information rate on each resource block, and defines the fairness index: $F = \left(\sum_{n=1}^{N} R_n\right)^2 \Big/ N \sum_{n=1}^{N} R_n^2$, where N is the total number of resource blocks in the system, and is the information rate on the resource block. The simulation parameters are shown in Table 1.

Table 1. Simulation parameters and values

Parameters	Values
System bandwidth	2.5 MHz
Total number of resource blocks	13
Path loss between D2D users	102.9 + 18.7log[d(km)]
User-to-relay path loss	103.8 + 20.9log[d(km)]
Path loss from relay to base station	100.7 + 23.5log[d(km)]
Standard deviation of shadow fading among D2D users	3 dB
Standard deviation of shadow fading from user to relay	10 dB
Standard deviation of shadow fading from relay to base station	6 dB
Relay transmission power range	20–30 dBm
User transmit power range	13–23 dBm
Maximum distance between D2D direct users	20 m
Relay coverage radius	200 m
Distance between base station and relay	125 m
Noise power spectral density	−174 dBm/Hz
Interference threshold	−70 dBm

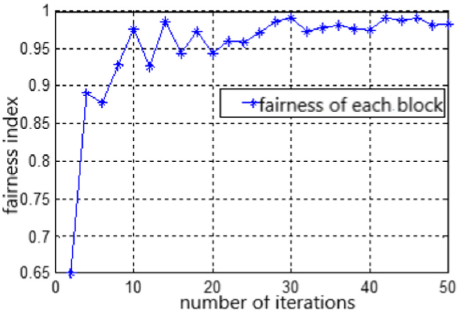

Fig. 7. Fairness index of each resource block

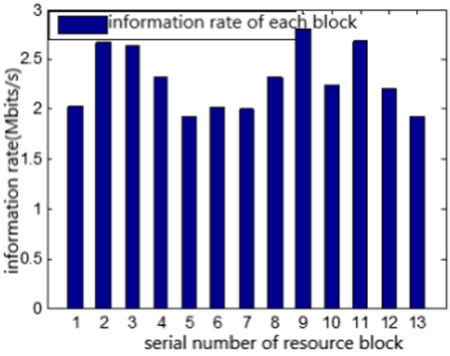

Fig. 8. Information rate distribution on each resource block

In order to avoid the inaccuracy of the final result caused by the randomness of the channel coefficient and the interference link coefficient, multiple simulations are used to average the result data to obtain the fairness index curve of the information rate distribution on each resource block. When the total system bandwidth and the total number of resource blocks are fixed, only the case of two relays is considered and the number of D2D user pairs and the number of cellular users under each relay are also the same. Still taking the number of D2D non-direcct user pairs and the number of cellular users under Relay 1 and Relay 2 is equal to 4 as an example, as shown in Fig. 7. The number of iterations gradually increases, and the fairness index of the information rate distribution of each resource block is gradually close to 1, which means the fairness becomes better. In the 50th iteration, the information rate distribution of each resource block is shown in Fig. 8, and the information rate on each resource block is distributed between 2 and 3 Mbit/s.

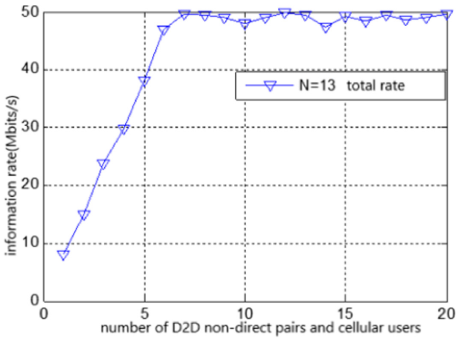

Fig. 9. Total system information rate

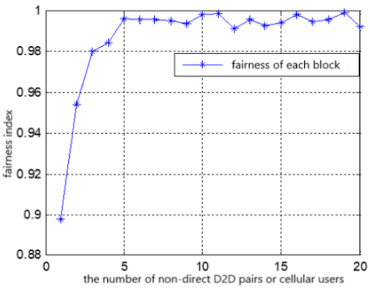

Fig. 10. Fairness of the information rate of each resource block

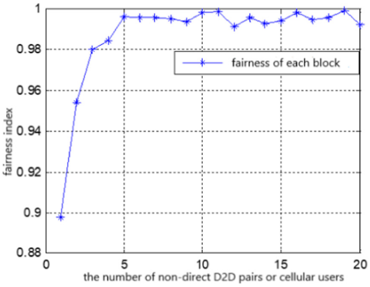

Fig. 11. Fairness of information rate of different types of users

As shown in Fig. 7 and Fig. 8, the number of iterations is 50, and the total number of resource blocks is 13. The simulation results show that as the number of D2D non-direct users and the number of cellular users under each relay increase, the total throughput of the system also first undergoes an approximately linear growth stage, and finally stabilizes. When the number of D2D non-direct users and the number of cellular users are greater than 7, the information rate reaches about 50 Mbit/s.

Next, observe the fairness of the information rate on each resource block under the simulation conditions in Fig. 9, as shown in Fig. 10 and Fig. 11, it can be seen that as the D2D pass-through user pairs under each relay As the number of users and the number of cellular users increase, the fairness index of the information rate on each resource block also increases, and gradually approaches 1. With the increase in the number of D2D direct users and the number of cellular users under each relay, the fairness of the information rate of the two shows a decreasing trend, and the greater the number of users, the worse the fairness.

5 Conclusion

This paper mainly studies the LTE-A cell throughput optimization problem that combines relay technology and D2D technology. It mainly includes two parts. The first part focuses on the actual cell system communication model, LTE-A relay cellular network architecture, link model and usage scenarios, as well as D2D communication principles and user working modes to construct a mathematical model of non-direct D2D user types. The second part uses the Lagrange multiplier method and KKT conditions to solve the constructed optimization problem with throughput as the optimization objective function.

This subject mainly completed the following work content:

(1) Under the LTE-A cellular network, combining relay technology, D2D technology and practical application scenarios together, a single cell model is constructed according to the working mode of non-direct D2D communication users. Besides, we analyze each communication process and interference

source to establish a MINLP optimization problem with power, interference, and user QoS requirements as constraints, system throughput as an objective function, user and relay transmission power, and user distribution factors on different resource blocks as optimization variables.

(2) From a mathematical point of view, the concavity and convexity of the optimization problem is proved. The result shows that the optimization problem is convex. Using the Lagrange multiplier method and the KKT conditions in the convex optimization theory, the optimal solution expression is derived. Through simulation analysis, it is found that the maximum throughput of the cellular system where D2D users are located in different working modes is very different. D2D direct users can bring greater throughput to the system than D2D non-direct users, but it reduces fairness of information rate distribution on resource blocks.

There are still many shortcomings in this article. For example, in the cell built on the basis of LTE-A cellular network, combined with relay technology and D2D technology, only power, user service quality QoS, and throughput limitations are considered. In the follow-up research process, performance indicators such as spectrum efficiency and energy efficiency can be added to continue to improve this model.

References

1. Khosrowpour, M.: An overview of 3GPP long term evolution (LTE). 13–19 (2015)
2. Phunchongharn, P., Hossain, E., Kim, D.I.: Resource allocation for device-to-device communications underlaying LTE-advanced networks. IEEE Wirel. Commun. **20**(4), 91–100 (2013)
3. Periyalwar, S., Zhang, H., Senarath, N., et al.: System and method for peer-to-peer communication in cellular systems. US, US 8260337 B2, pp. 51–60 (2012)
4. Park, K.J., Cho, H.G., Kwon, Y.H., et al.: Method for transmitting an uplink signal and feedback information, and relay apparatus using the method. US, US 20140286265 A1, pp. 56–58 (2014)
5. Zhu, K., Hossain, E.: Joint mode selection and spectrum partitioning for device-to-device communication: a dynamic Stackelberg game. IEEE Trans. Wirel. Commun. **14**(3), 1406–1420 (2015)
6. Ghaboosi, K., Vemulapalli, M.G., Oates, N.J.: Duplexing in long term evolution cellular networks. US, US20140286205, pp. 21–31 (2014)
7. Cha, W.S., Song, J.K., Kim, S.T., et al.: Method of selecting relay mode in mobile ad-hoc network. US9060386, pp. 54–68 (2015)
8. Lee, K., Lee, J., Yi, Y., et al.: mobile data offloading: how much can WiFi deliver. IEEE/ACM Trans. Netw. **21**(2), 536–550 (2013)
9. Ristanovic, N., Boudec, J.Y.L., Chaintreau, A., et al.: Energy efficient offloading of 3G networks, pp. 202–211 (2011)
10. Joung, J., Sun, S.: Power efficient resource allocation for downlink OFDMA relay cellular networks. IEEE Trans. Signal Process. **60**(5), 2447–2459 (2012)
11. Alam, M.S., Mark, J.W., Shen, X.: Relay selection and resource allocation for multi-user cooperative LTE-A uplink. In: IEEE International Conference on Communications, pp. 5092–5096. IEEE (2012)

12. Choi, Y., Ji, H.W., Park, J.Y., et al.: A 3W network strategy for mobile data traffic offloading. IEEE Commun. Mag. **49**(10), 118–123 (2011)
13. Korhonen, J., Savolainen, T., Ding, A.Y., et al.: Toward network controlled IP traffic offloading. IEEE Commun. Mag. **51**(51), 96–102 (2013)
14. Doppler, C., Ribeiro, B.: Mode selection for device-to-device communication underlaying an LTE-advanced network. IEEE Xplore **29**(16), 1–6 (2010)
15. Yu, C.H., Doppler, K., Ribeiro, C.B., et al.: Resource sharing optimization for device-to-device communication underlaying cellular networks. IEEE Trans. Wirel. Commun. **10**(8), 2752–2763 (2011)
16. Kaufman, B., Lilleberg, J., Aazhang, B.: Spectrum sharing scheme between cellular users and ad-hoc device-to-device users. IEEE Trans. Wirel. Commun. **12**(3), 1038–1049 (2013)
17. Min, H., Seo, W., Lee, J., et al.: Reliability improvement using receive mode selection in the device-to-device uplink period underlaying cellular networks. IEEE Trans. Wirel. Commun. **10**(2), 413–418 (2011)
18. Dong, H.L., Choi, K.W., Jeon, W.S., et al.: Two-stage semi-distributed resource management for device-to-device communication in cellular networks. IEEE Trans. Wirel. Commun. **13**(4), 1908–1920 (2014)
19. Tehrani, M.N., Uysal, M., Yanikomeroglu, H.: Device-to-device communication in 5G cellular networks: challenges, solutions, and future directions. IEEE Commun. Mag. **52**(5), 86–92 (2014)
20. Li, Y., Wang, Z., Jin, D., et al.: Optimal mobile content downloading in device-to-device communication underlaying cellular networks. IEEE Trans. Wirel. Commun. **13**(7), 3596–3608 (2014)